Firms, Informality and Development: Theory and evidence from Brazil

Gabriel Ulyssea*

Online Appendix

A Cross country evidence

Figure A.1's left panel displays informal sector's size in Latin American countries, which is measured as the share of employees not covered by social security.[1] The right panel shows the c.d.f. plot of informal sector's size for 116 countries that have a GDP per capita that is less or equal to half of USA's. The size measure used in this graph is informal sector's share of GDP, which comes from La Porta and Shleifer (2008).

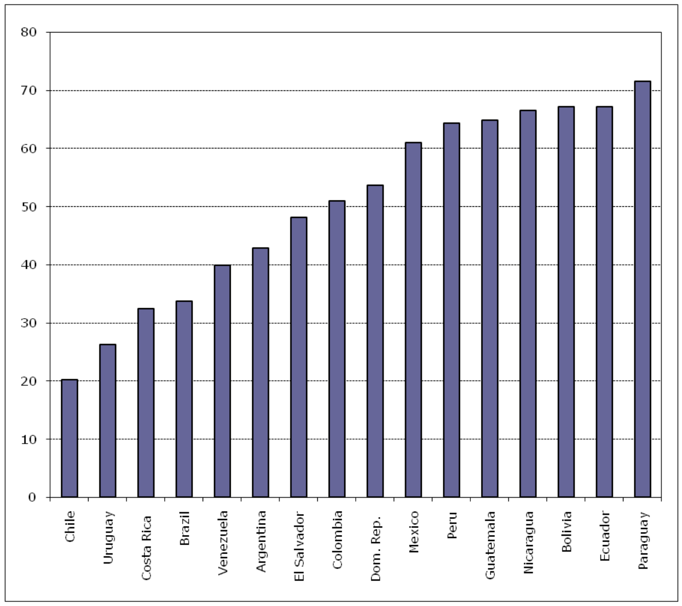
Table A.1: Regulatory costs

	Entry Costs		**Labor tax (%)**
	# Procedures	# Days	
E. Asia & Pacific	7	37	10.7
E. Europe & C. Asia	6	16	21.7
L. A.C.	9	54	14.6
Mid. East & N. Africa	8	20	16.9
OECD high income	5	12	24
South Asia	7	23	7.7
Sub-Saharan Africa	8	37	13.5
Brazil	**13**	**119**	**40.9**

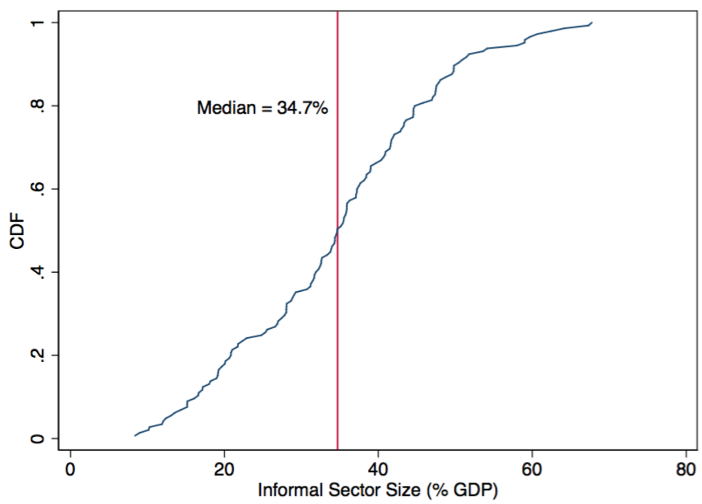
Source: Doing Business Initiative, 2010 (*www.doingbusiness.org*).

*Department of Economics, University of Oxford, Manor Road Building, Manor Road, Oxford, OX1 3UQ (email: gabriel.ulyssea@economics.ox.ac.uk).

[1]The data come from the Socio-Economic Data Base for Latin America and the Caribbean (SEDLAC), a joint initiative by the World Bank and Universidad Nacional de La Plata (available at *http://sedlac.econo.unlp.edu.ar/esp/*).



(a) Labor informality (Latin America)



(b) CDF of size measure (developing countries)

Figure A.1: Informal sector's size

B Data appendix

As described in Section II.A, the two main data sets used in this paper are the ECINF survey (*Pesquisa de Economia Informal Urbana*) and the *Relacao Anual de Informacoes Sociais* (RAIS), an administrative data set from the Brazilian Ministry of Labor. In both data sets, I exclude the public sector and agriculture. In RAIS, I exclude firms that declare a wage bill equal to zero. Since the RAIS data set contains the universe of formal firms, I use a 25% random sample from the original data set to decrease the computational burden.

As for the ECINF, some additional filters were applied. Many of the observations regard self employed individuals, street vendors and other activities that do not correspond to the standard definition of a firm. In order to obtain the most comparable unit of analysis with the formal firms covered by the RAIS data set, I dropped the entrepreneurs who declared to have another job, and who do not have a specific physical location outside their household where their activity takes place. To avoid outliers, I trimmed the first and 99th percentiles of log-revenues distribution.

C Additional Facts

Table C.1: Descriptive Statistics – Workers

	Informal		Formal	
	Mean	SD	Mean	SD
Log(wage)	6.236	0.761	6.735	0.683
Share High Skill	0.355	0.479	0.504	0.500
Sectoral Composition				
Retail	0.248	0.432	0.236	0.425
Construction	0.144	0.351	0.052	0.221
Manufacturing	0.174	0.379	0.278	0.448
Services	0.213	0.409	0.156	0.363
Age	31.8	11.4	33.4	10.4
Male	0.650	0.477	0.632	0.482

Notes: Data from National Household Survey (PNAD). Informal employees are defined as indicated in the text: those employees who do not hold a formal labor contract, which in Brazil is defined by having a booklet (*carteira de trabalho*). High skilled workers are those with at least a high school degree.

Table C.2: Formal and informal employment composition by firm size

	Informal Workers (in %)	Formal workers (in %)
Firm size (# employees)		
0–5	35.8	6.6
6–10	11.7	7.2
11 or more	52.5	86.2

Source: Author's own tabulations from the Monthly Employment Survey (PME) 2003.

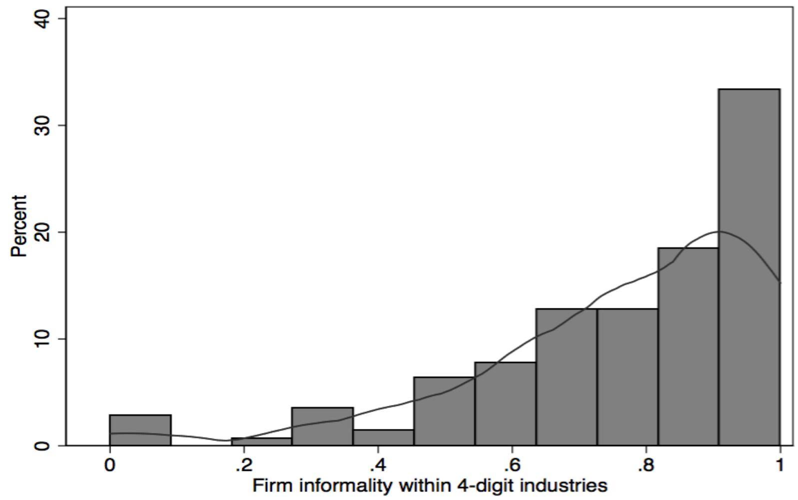


Figure C.1: Share of informal firms at the 4-digit industry level: Histogram

Note: The variable used is the share of informal firms measured at the 5-digit industry level. The figure shows the histogram of this industry-specific measure of firm informality.

D Model with Homogeneous Workers

D.1 The cost function $\tau_i(\ell)$ and the detection probability

In this Subsection I argue that the cost function $\tau_i(\cdot)$ can be directly obtained from a formulation that explicitly accounts for a detection probability. In particular, one could use an alternative formulation for the profit function described in Section III.A, as follows:

$$\Pi_i(\theta) = \max_{l} \left\{ \left[1 - p(\ell)\right] \theta q(\ell) - w\ell \right\}$$

where $0 < p(\cdot) \leq 1$ is strictly increasing and convex.

The $p(\ell)$ can be interpreted as the probability of being caught by government's officials, in which case the informal firm looses all of its production.[2] To obtain a direct correspondence between this formulation and expression **??**, one could parametrize the labor distortion as $\tau_i(\ell) = \frac{1}{1-p(\ell)}$, so that as $\lim_{p(\ell)\to 1} \tau_i(\ell) = \infty$ and $\lim_{p(\ell)\to 0} \tau_i(\ell) = \ell$.

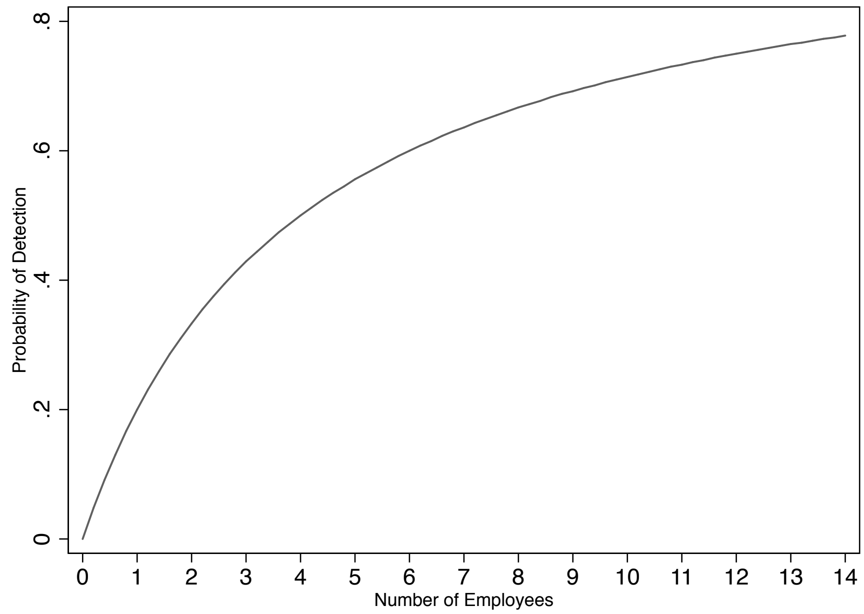
Figure D.1 shows the probability of detection $p(\ell)$ that corresponds to the function $\tau_i(\ell) = \left(1 + \frac{\ell}{b_i}\right)$ using the estimated value of b_i as displayed in Table **??**.

D.2 Productivity distributions in both sectors

The post-entry, unconditional productivity distribution in the informal and formal sectors, respectively, is given by the following expressions:

[2] Alternatively, it can be thought as a probability of detection combined with a fine that is proportional to firm's revenues.

Figure D.1: The detection probability that corresponds to the cost function $\tau_i(\ell)$



Note: The figure shows, for each firm's size (measured as number of employees), the value of the probability of detection given by $\tau_i(\ell) = \frac{\ell}{1-p(\ell)}$.

$$
\begin{aligned}
f_{\theta_i}(x) &= \frac{1}{G(\overline{\nu}_f) - G(\overline{\nu}_i)} \int_{\overline{\nu}_i}^{\overline{\nu}_f} f(x|\nu)\, dG(\nu) \\
f_{\theta_f}(x) &= \frac{1}{1 - G(\overline{\nu}_f)} \int_{\overline{\nu}_f}^{\infty} f(x|\nu)\, dG(\nu)
\end{aligned}
\tag{1}
$$

where f_{θ_s} is absolutely continuous and $F_{\theta_s}(\cdot)$ denotes the corresponding c.d.f..

As mentioned above, firms can be surprised with a bad productivity draw. Those with a $\theta < \overline{\theta}_s$, where $\overline{\theta}_s$ is such that $\pi_s\left(\overline{\theta}_s, w\right) = 0$, will not produce and will leave immediately. Hence, the effective productivity distribution among successful entrants is given by the following expressions:

$$
\tilde{f}_{\theta_s}(x) = \begin{cases} \frac{f_{\theta_s}(x)}{1-F_{\theta_s}(\overline{\theta}_s)} & \text{if} \quad x \geq \overline{\theta}_s \\ 0 & \text{if} \quad \theta < \overline{\theta}_s \end{cases} \tag{2}
$$

where $s = i, f$.

D.3 Uniqueness of Equilibrium

This section contains a simple argument to prove the uniqueness of equilibrium. The key equilibrium conditions are given by the zero profit conditions, the free entry conditions and the market clearing condition, respectively:

$$\begin{aligned}\pi_s\left(\bar{\theta}_s, w\right) &\equiv \Pi_s\left(\bar{\theta}_s, w\right) - \bar{c} = 0 &(3)\\ V_i^e\left(\overline{\nu}_i, w\right) &= E_i &(4)\\ V_f^e\left(\overline{\nu}_f, w\right) &= V_i^e\left(\overline{\nu}_f, w\right) - \left(E_i - E_f\right) &(5)\\ L_i + L_f &= \bar{L} &(6)\end{aligned}$$

where $s = i, f$ and the free entry conditions assume that entry is positive in both sectors.

Fix a given wage. Given the assumptions made for the cost and production functions, the functions $\pi_s(\theta, w)$ are strictly increasing in θ and decreasing in w. Moreover, as $\bar{c} > 0$, there is a $\theta > 0$ such $\pi_s(\theta, w) < 0$. Thus, there is a unique $\bar{\theta}_s$ such that 3 holds. The simple form of the value functions, $V_s(\theta, w) = \max\left\{0, \frac{\pi_s(\theta,w)}{\kappa_s}\right\}$, implies that they are also continuous and strictly increasing in θ and decreasing in w. Combining this last fact with the assumptions made about $F(\theta|\nu)$, it follows that that there is an unique $\bar{\nu}_s$, $s = i, f$, such that free entry conditions hold, and that $\bar{\nu}_f > \bar{\nu}_i$. The latter follows from the assumption that $E_f > E_i$. The unique entry thresholds pin down the mass of entrants in both sectors: $M_i = \left[G\left(\overline{\nu}_f\right) - G\left(\overline{\nu}_i\right)\right]M$ and $M_f = \left[1 - G\left(\overline{\nu}_f\right)\right]M$. Given the mass of entrants in each sector, and the unique thresholds $\bar{\theta}_s$, the flow conditions in both sectors [given by (**??**)] pin down the mass of firms in each sector, μ_s. The last condition to close the equilibrium determination is the market clearing condition for the labor market, which determines the equilibrium wage. Of course, if $L^d \equiv L_i + L_f > \bar{L}$ there is excess demand and the wage will increase up until the point where $L^d = \bar{L}$ (the symmetric argument is true for excess supply). Because of the properties of the profit functions, the individual labor demand functions $\ell^*(\theta, w)$ are also continuous, single valued, and strictly increasing in θ and decreasing in w. Thus, there is an unique wage such that $L^d = \bar{L}$.

E Model with Heterogeneous Workers

Informal firm's problem can be solved in two steps. The first solves the following cost minimization problem:

$$\begin{aligned}\min_{l_1, l_2} \quad & w_1 l_1 + w_2 l_2\\ \text{s.t. } \ell_i &= \left[\eta_i l_1^{\rho} + (1 - \eta_i)\, l_2^{\rho}\right]^{\frac{1}{\rho}}\end{aligned}$$

where the term $\tau_i(\ell_i)$ is suppressed, as it is fixed once ℓ_i is fixed and it affects both factors equally.

The first order conditions imply

$$\frac{l_1}{l_2} = \left(\frac{(1 - \eta_i)\, w_1}{\eta_i w_2}\right)^{\frac{1}{\rho - 1}} \tag{7}$$

and substituting into the production function and solving for l_1 and l_2, one obtains

$$l_1 = \left(\frac{w_1}{w_i}\right)^{\frac{1}{\rho-1}} \left(\frac{\ell_i}{\eta_i^{\frac{1}{\rho-1}}}\right) \tag{8}$$

$$l_2 = \left(\frac{w_2}{w_i}\right)^{\frac{1}{\rho-1}} \left(\frac{\ell_i}{(1-\eta_i)^{\frac{1}{\rho-1}}}\right) \tag{9}$$

where $w_i = \left[\eta_i \left(\frac{w_1}{\eta_i}\right)^{\frac{\rho}{\rho-1}} + (1-\eta_i)\left(\frac{w_2}{(1-\eta_i)}\right)^{\frac{\rho}{\rho-1}}\right]^{\frac{\rho-1}{\rho}}$.

Using the expressions for l_1 and l_2 one obtains the cost function for a given level of output ℓ_i:

$$c(\ell_i) = \tau_i(\ell_i)\,\ell_i w_i$$

Thus, one can write informal firm's problem as

$$\max_{\ell_i} \theta(\ell_i)^\alpha - \tau_i(\ell_i)\,\ell_i w_i$$

The above problem pins down the ℓ_i, while the l_1 and l_2 are given by expressions (8) and (9), respectively.

Formal Firms

As presented in Section III, formal firm's profit maximization problem can be written as follows:

$$\Pi_f(\theta, \mathbf{w}) = \max_{l_1, l_2} \left\{(1-\tau_y)\,\theta\ell^\alpha - C(l_1, l_2)\right\} \tag{10}$$

and

$$C(l_1, l_2) = \begin{cases} \tau_{f1}(l_1)\,w_1 + \tau_{f2}(l_2)\,w_2, \text{ for } l_s \leq \tilde{l}_s, s = 1,2 \\ \tau_{f1}\left(\tilde{l}_1\right) w_1 + (1+\tau_w)\,w_1\left(l_1 - \tilde{l}_1\right) + \tau_{f2}(l_2)\,w_2, \text{ for } l_1 > \tilde{l}_1, l_2 \leq \tilde{l}_2 \\ \sum_{s=1,2}\left\{\tau_{fs}\left(\tilde{l}_s\right) w_s + (1+\tau_w)\,w_s\left(l_s - \tilde{l}_s\right)\right\}, \text{ for } l_s > \tilde{l}_s, s = 1,2 \end{cases}$$

The problem for the firm that only hires informal workers is analogous to the informal firm. For formal firms that hire some formal workers for both types, one can write:

$$\max_{l_{1f}, l_{2f}} (1-\tau_y)\theta\ell_f^\alpha - (1+\tau_w)\left(w_1 l_{1f} + w_2 l_{2f}\right) - \textstyle\sum_{s=1,2} \tau_{fs}\left(\tilde{l}_s\right) w_s$$
$$\text{s.t.} \quad \ell_f = \left[\eta_f\left(l_{1f} + \tilde{l}_1\right)^\rho + (1-\eta_f)\left(l_{2f} + \tilde{l}_2\right)^\rho\right]^{\frac{1}{\rho}}$$

Again, one can solve this problem by solving the cost minimization problem first:

$$\begin{aligned} \min_{l_{1f},l_{2f}} \quad & w_1 l_{1f} + w_2 l_{2f} \\ \text{s.t.} \quad & \ell_f = \left[\eta_f \left(l_{1f} + \tilde{l}_1\right)^\rho + (1-\eta_f)\left(l_{2f} + \tilde{l}_2\right)^\rho\right]^{\frac{1}{\rho}} \end{aligned}$$

where the payroll tax is suppressed as it affects both factors equally.

From the FOC:

$$\frac{\eta_f}{1-\eta_f}\left(\frac{\tilde{l}_1 + l_1}{\tilde{l}_2 + l_2}\right)^{\rho-1} = \frac{w_1}{w_2} \tag{11}$$

Using this relationship and solving for l_{1f} and substituting back into the production function, one can obtain the following expressions:

$$l_{1f} = \left(\frac{w_1}{w_f}\right)^{\frac{1}{\rho-1}} \frac{\ell_f}{\eta_f} - \tilde{l}_1 \tag{12}$$

$$l_{2f} = \left(\frac{w_2}{w_f}\right)^{\frac{1}{\rho-1}} \frac{\ell_f}{(1-\eta_f)} - \tilde{l}_2 \tag{13}$$

where $w_f = \left[\eta_f \left(\frac{w_1}{\eta_f}\right)^{\frac{\rho}{\rho-1}} + (1-\eta_f)\left(\frac{w_2}{(1-\eta_f)}\right)^{\frac{\rho}{\rho-1}}\right]^{\frac{\rho-1}{\rho}}$.

The profit maximization thus simplifies to:

$$\max_{\ell_f} (1-\tau_y)\theta\ell_f^\alpha - (1+\tau_w)\,\ell_f w_f$$

where l_{1f} e l_{2f} are given by (12) and (13), respectively.

F Estimation Appendix

F.1 Implementation Details

For the estimation, I consider $S = 20$ simulated data sets containing a mass of $M = 300,000$ potential entrants each. For each potential entrant, I draw a pre-entry productivity parameter (ν) and a post entry productivity shock (ϵ). The stochastic components of the model are drawn only once in the beginning of the procedure and are kept fixed during the estimation.

The estimation procedure is done conditional on the observed wages for low and high skill workers. In order to purge the variation that is due to differences in observables, I estimate a log-wage regression with a dummy for high skill workers (at least completed high school), dummies for male, 4-digit industry classification (164 dummies), state of residence (26 dummies), holds a formal contract, white, age and age squared, tenure in current job and tenure squared. To further minimize measurement error, I restrict the sample to employees only (formal or informal), who are 18 to 69 years old, and who have worked at least 20 hours in the reference week but at most 84 hours (which is the 99th percentile). The lower bound of 20 hours aims at excluding interns who are still in school and workers with very low attachment to the labor market. I use the estimated coefficients to compute the adjusted wage for low and high skill workers. Of course, the goal is not to recover the returns to schooling but to obtain wage measures for low and high skill workers that net out variation from workers' observable characteristics and which are more compatible with wages in the theoretical model.

Since each potential entrant has an individual pre-entry productivity parameter, it is necessary to compute the expected value of entry conditional on each individual parameter. To save on computational time, I use 111 equally spaced grid points for the productivity space, where the maximum value in the productivity grid implies a firm size of more than 18,000 employees and is not binding. I compute a vector of transition probabilities for each point in the grid in order to compute the expected post-entry values in each sector for each potential entrant. For that, I use the method proposed by Tauchen (1986).

F.1.1 Smoothing the Policy Functions

One difficulty when estimating discrete choice models using simulation-based methods is that simulated choices, such as the decision to enter the formal sector, will be a step function of the parameter vector (φ) given the random draws (ν and ε). Since these discontinuities are inherited by the objective function, this also precludes the use of derivative-based methods, which are faster and more accurate than derivative-free methods or random search algorithms (Bruins et al., 2015). To overcome these challenges, I use the following smoothing function proposed by Bruins et al. (2015) to correct for the choppiness of the policy functions (see also Keane and Smith, 2003):

$$h\left(\tilde{V}(\varphi), m, \lambda\right) = \frac{\tilde{V}_m(\varphi)/\lambda}{1 + \sum_k \tilde{V}_k(\varphi)/\lambda}$$

where $\tilde{V}(\varphi)$ is the set of net payoffs attached to the choices firms have. For example, potential entrants can either enter the formal or informal sectors, or not enter at all; for those that enter, the choices are to remain in their sector or to exit immediately. $\tilde{V}_m(\varphi)$ denotes the net payoff of the specific choice m (e.g. to enter the formal sector) and λ is the smoothing parameter.

As the smoothing parameter λ goes to zero, $h(\cdot)$ goes to one if alternative m provides the highest payoff, and zero otherwise. When choosing λ, one must consider two opposing forces: bias and smoothness. Large values of λ are better to smooth the objective function but can lead to biased estimates. Small values of λ reduce this bias but increase the choppiness of the objective function (Keane and Smith, 2003). This latter effect can be alleviated by choosing a large number of simulated data sets. As in Altonji et al. (2013), I choose $\lambda = 0.05$ and, as mentioned above, $S = 20$.

F.2 Standard Errors

The estimator is given by

$$\hat{\varphi} \quad = \quad \arg\min_{\varphi} Q\left(\varphi\right) = \left\{ g_{NS}\left(\varphi\right)' \hat{\mathbf{W}} g_{NS}\left(\varphi\right) \right\}$$

where, as discussed in the text, $g_{NS}\left(\varphi\right) = \hat{m}_N - \tilde{m}_S(\varphi)$, and I omit the conditioning arguments for notational convenience.

The conditions for consistency are close to the ones for extremum estimators. In fact, the substantive assumptions are exactly the same as in the GMM case, namely, the GMM identification assumption, and the requirement that the parameter space is compact (see Newey and McFadden, 1994). The main difference thus lies on the type of regularity conditions required to guarantee consistency. I follow the discussion in Duffie and Singleton (1993), who provide conditions for both weak and strong consistency, which are satisfied by the present model.[3] The following assumptions are made for asymptotic normality to hold: (i) φ_0 and $\hat{\varphi}$ are interior to the parameter space; (ii) the simulator used to generate the simulated data is continuously differentiable w.r.t. φ in a neighborhood $\mathcal{B}$ of φ_0; and (iii) $\mathbf{G}_0 \equiv E\left[\nabla_{\varphi} g_{NS}\left(\varphi_0\right)\right]$ exists, is finite and $\mathbf{G}_0'\mathbf{W}\mathbf{G}_0$ is nonsingular, where $\mathbf{W}$ is a positive semi-definite matrix.

The derivation of the asymptotic distribution is standard (e.g. Gourieroux and Monfort, 1996), which gives the following expression:

$$\sqrt{N}\left(\hat{\varphi}-\varphi_0\right) \xrightarrow{d} N\left(0,\left(\mathbf{G}_0'\mathbf{W}\mathbf{G}_0\right)^{-1}\mathbf{G}_0'\mathbf{W}\Sigma_s\mathbf{W}\mathbf{G}_0\left(\mathbf{G}_0'\mathbf{W}\mathbf{G}_0\right)^{-1}\right)$$

where $\Sigma_s = \zeta\Sigma$, $\zeta = \lim_{N\to\infty}\left(1+\frac{N}{L}\right)$ (Gourinchas and Parker, 2002); N is the number of observations used to obtain the vector of moments, and L is the number of simulations.

The choice of the optimal weighting matrix implies that $\hat{\mathbf{W}}_N \longrightarrow \Sigma$, where $\Sigma = E\left[g\left(\varphi_0\right)g\left(\varphi_0\right)'\right]$ is the GMM asymptotic variance-covariance matrix. Analogous to the GMM estimator, the optimal weighting matrix is given by $W^* = \Sigma_s^{-1}$, which therefore reduces the asymptotic variance-covariance matrix to

$$V_s\left(W^*\right) = \left(\mathbf{G}_0'\mathbf{\Sigma_s^{-1}}\mathbf{G}_0\right)^{-1}$$

The actual variance-covariance is computed using the empirical counterpart of Σ, a diagonal matrix with the empirical variances of the moments on the diagonal. Some of the moments are obtained from simple means (such as the share of informal workers) and some are obtained from regressions, which is a simple way to compute conditional means (such as the share of informal firms by firm size). Following Adda et al. (2017), for the moments that are based on regressions I use robust standard errors and for the moments that consist of means I compute their variance using a bootstrap method with 500 replications. I have experimented with bootstrapping some of the moments obtained through regressions and the estimated standard errors are essentially the same, presenting differences only in the fourth or fifth decimal place. For G_0, I use its computational equivalent, which can be obtained using standard numerical differentiation methods.

F.3 Identification

To further investigate whether the chosen moments allow me to identify all the parameters, I follow the analysis in Adda et al. (2017). The basic motivation is that if the model is well-identified, the objective function should not be flat in the region around the vector of estimated parameters. If the objective function is flat, then this is would raise identification concerns, as this could reflect that, for example, the moments chosen are irrelevant to identify the parameters of the model. To asses that, I compute the objective function for each parameter at a value 1, 2 and 5 percent away from its estimated value and compare it to the objective function evaluate at the estimated vector. The idea is to evaluate how convex the objective function is at the estimated vector of parameters.

[3] Pakes and Polard (1989) provide the regularity conditions under which a broad class of simulation-based estimators are both consistent and asymptotically normal,which includes the SMM estimator proposed by McFadden (1989).

As in Adda et al. (2017), I plot the percentage change in the objective function from the percentage change in each of the parameters. Figure F.1 shows the results. As the figure shows, there is enough variation in the objective function, which is reassuring of the model's identification.

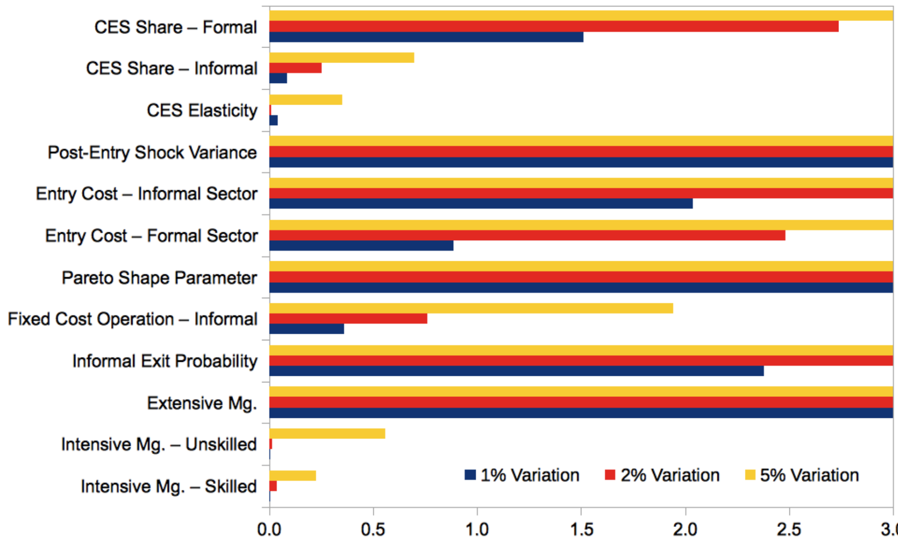


Figure F.1: Sensitivity of the Objective Function

Note: The horizontal bars show the percentage change in the objective function with respect to one, two and five percent changes in the given parameter of the model. The figure is truncated on the right at 3%.

References

Adda, J., C. Dustmann, and K. Stevens (2017). The career costs of children. *Journal of Political Economy 125*(2), 293–337.

Altonji, J. G., A. A. Smith, and I. Vidangos (2013). Modeling earnings dynamics. *Econometrica 81*(4), 1395–1454.

Bruins, M., J. A. Duffy, M. P. Keane, and A. A. Smith Jr (2015). Generalized indirect inference for discrete choice models. Yale University, mimeo.

Duffie, D. and K. Singleton (1993). Simulated moments estimation of markov models of asset prices. *Econometrica 61*(4), 929–952.

Gourieroux, C. and A. Monfort (1996). *Simulation-Based Econometric Methods*. Oxford University Press.

Gourinchas, P.-O. and J. Parker (2002). Consumption over the life cycle. *Econometrica 70*(1), 47–89.

Keane, M. and A. A. Smith (2003). Generalized indirect inference for discrete choice models. Yale University, mimeo.

La Porta, R. and A. Shleifer (2008). The unofficial economy and economic development. *Brooking Papers on Economic Activity 105*(3), 473–522.

McFadden, D. (1989). A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica 57*(5), 995–1026.

Newey, W. K. and D. McFadden (1994). Large sample estimation and hypothesis testing. In R. F. Engle and D. L. McFadden (Eds.), *Handbook of Econometrics*, Volume 4 of *Handbook of Econometrics*, pp. 2111 – 2245. Elsevier.

Pakes, A. and D. Polard (1989). Simulation and the asymptotics of optimization estimators. *Econometrica 57*(5), 1027–1057.

Tauchen, G. (1986). Finite state markov-chain approximation to univariate and vector autoregressions. *Economic Letters 20*, 177–81.